Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Volume I: Foundations

Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics Volume II: Applications

Editors

Krassimir T. Atanassov
Michał Baczyński
Józef Drewniak
Janusz Kacprzyk
Maciej Krawczak
Eulalia Szmidt
Maciej Wygralak
Sławomir Zadrożny

Editors

Krassimir T. Atanassov
Władysław Homenda
Olgierd Hryniewicz
Janusz Kacprzyk
Maciej Krawczak
Zbigniew Nahorski
Eulalia Szmidt
Sławomir Zadrożny

SRI PAS IBS PAN

Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics
Volume II: Applications

Systems Research Institute
Polish Academy of Sciences

Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics
Volume II: Applications

Editors

Krassimir T. Atanassov
Władysław Homenda
Olgierd Hryniewicz
Janusz Kacprzyk
Maciej Krawczak
Zbigniew Nahorski
Eulalia Szmidt
Sławomir Zadrożny

IBS PAN SRI PAS

Systems Research Institute
Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
www.ibspan.waw.pl

ISBN 9788389475305

On linguistic interpretation of data-driven knowledge

Ewa Straszecka
Institute of Electronics, Silesian University of Technology
16 Akademicka St., 44-100 Gliwice, Poland
ewa.straszecka@polsl.pl

Abstract

The paper deals with the problem of interpretation of linguistic values in heuristic rules. Such rules are often made by experts in order to support a diagnosis. However, membership functions created by the experts to represent premises of the rules as well as weights of symptoms in the diagnosis may strongly depend on diagnostic circumstances. Therefore, the goal of the proposed method is to maintain rules and to attach them with membership functions and weights that are data-driven. In this way the rules preserve their semantic meaning and if some of them are not applicable they are deleted.

Keywords: Heuristic rules, diagnosis support, the Dempster-Shafer theory, fuzzy sets.

1 Introduction

Diagnosis support systems that are based on rules which represent heuristics are generally better accepted than methods which do not offer semantic interpretation of knowledge [4], [5]. However, it is not always possible to create the rules during a dialog with an expert because of time and cost limitations. Moreover, knowledge should be continuously updated which imply prevalence of diagnosis support based on training data included in databases, rather than on expert's

Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Volume II: Applications *(K.T. Atanassow, W. Homenda, O. Hryniewicz, J. Kacprzyk, M. Krawczak, Z. Nahorski E. Szmidt, S. Zadrożny, Eds.), IBS PAN - SRI PAS, Warsaw, 2009.*

consultation. However, data-driven rules often lack semantic meaning. Thus, it should be considered if it is possible to sustain rules which are made by an expert and provide them with changing representation of their linguistic values. This is not very difficult in case of designing membership functions according to various training data. Still, weights of the rules should be changed along with a modification of the functions. Sometimes a weight should even be equal to zero, so that a rule is excluded from reasoning. The paper suggests a method of updating rules in the framework of the Dempster-Shafer theory [3] extended for fuzzy focal elements [6]. In this method symptoms are focal elements that are represented by fuzzy sets. A focal element may include one or several symptoms. The focal elements imply a diagnosis, thus rules 'IF symptom(s) THEN diagnosis' are constructed. Each rule is assessed with a weight which is the value of the basic probability assignment [3] for the focal element. Membership functions of the fuzzy sets and the weights are determined by means of training data. This makes it possible to use the same heuristics in different diagnostic circumstances.

2 Methods

A diagnostic rule made by an expert has the form:

$$\text{IF } X_1 \text{ is } A_{1r}, \text{ and } \ldots, \text{ and } X_n \text{ is } A_{nr} \text{ THEN } D_l \ (w_r), \tag{1}$$

where X_j is a symptom, A_{jr} – its linguistic value occurring in the $r-th$ rule, $j = 1, \ldots, n$; D_l is a diagnosis, $l = 1, \ldots, N$; and w_r is the rule weight in the diagnosis. The A_{jr} linguistic value is represented by the $\mu_r(x_j)$ membership function. This function is often obtained as the mean opinion of n_e experts, each of them constructing the $\mu_{rk}(x_j)$ function. The final function is assumed to be their average opinion $\mu_r(x_j)$:

$$\mu_r(x_j) = \frac{\sum_{k=1}^{n_e} \mu_{rk}(x_j)}{n_e}.$$

However, each expert's function instead of the mean estimation would be more appropriate if the experts work in different diagnostic environments. In such a case, the function should be created for individual training data. For instance, it can be made using the method described in [6] in which four crucial points of a trapezoidal membership function are determined by intersection points between probability distributions and by quartiles of training data. The basic shape can be next tuned by means of moving the two upper points of the trapezoidal function toward a triangle and the characteristic function. Let us denote by δ the coefficient by which the distance between quartiles are multiplied to determine the new position of the points. Thus, the function can be changed toward a triangular function (with negative δ coefficient) or toward an almost-characteristic function (with positive δ coefficient). The shape that results in the lowest classification error can be accepted for further training (see [6] for details of tuning). Still, simultaneously, the weight of the rule has to be determined for the same data. This is more difficult because w_r in (1) is usually a probability of D_l given $X_1, \ldots, X_n$ symptoms. If the classical definition of conditional probability is used, then dependence of symptoms has to be considered. This involves complex computations. Thus, the basic probability assignment (BPA) is proposed to be used, in which the only requirements are [3]:

$$m(f) = 0, \sum_{\alpha \in \mathbf{A}} m(\alpha) = 1, \tag{2}$$

where f stands for the false predicate and $\mathbf{A}$ is a set of focal elements α. In the diagnosis symptoms are focal elements. They can be single, i.e. concerning one symptom, or complex – that refer to several symptoms. The symptom represented by the $\mu_r(x_j)$ membership functions can be considered as the fuzzy focal element [6]. The fuzzy focal element may be presumed present if its membership is smaller than 1. Thus, a threshold should be determined which would evaluate presence of the symptom. Let us denote it η_{BPA}. This threshold is particularly important to decide if a symptom occurs in training data of some diagnosis. The normalized frequency of occurrence can be the BPA. Hence, in case of fuzzy focal elements the following condition:

$$\sum_{\substack{\alpha \in \mathbf{A} \\ \eta_\alpha > \eta_{BPA}}} m(\alpha) = 1, \tag{3}$$

holds true. The BPA (3) has to be calculated separately for each diagnosis. When α represents the premise of a diagnostic rule, than $w_r = m_l(\alpha)$. The latter value of the BPA can be found as frequency occurrence of a single or multiple symptoms included in the premise of the $r - th$ rule in training data set concerning $l - th$ diagnosis. The rules with their membership functions and BPAs create the knowledge base.

The diagnosis is the process in which evidence, i.e. observations, are matched with knowledge and beliefs in several diagnostic hypotheses are evaluated. Hence, an observation is matched with a membership function of an appropriate rule. Each symptom that is confirmed by the observation increase our belief in a diagnostic hypothesis. In case of a fuzzy symptom, i.e. a fuzzy focal element, the accuracy of matching can be lower than 1, so again a threshold should be assumed to decide whether the symptom is sufficiently confirmed. Therefore, the belief in the diagnostic hypothesis is given by an extended formula of belief measure originally proposed in the Dempster-Shafer theory [6]:

$$Bel(D_l) = \sum_{\substack{\alpha \in \mathbf{A} \\ \eta_\alpha > \eta_T}} m(\alpha), \tag{4}$$

where η_α denotes the minimal membership value that is indicated by observations matching all symptoms that refer to the α focal element. Hence, η_T is a threshold that is assumed for the diagnosis. The final diagnosis is determined after a comparison of belief values found for all diagnostic hypotheses [6]. This approach is very close to a classification, yet this is the classification according to criteria that have semantic meaning. The hypothesis with the greatest value is the final diagnosis.

Hence, both membership functions and weights of rules can be found by means of training data. Therefore, the same rules may be used with different membership functions and weights. In this way knowledge can be adopted to different diagnostic circumstances. Specifically, rules that concern symptoms that are not examined because of differences in diagnostic patterns can be flexibly removed. Moreover, it is possible to neglect rules related to variables that disturb classification. This can be done by means of the combination of fuzzy focal elements:

$$m_l(\alpha) = \frac{\sum_{\alpha=\beta\cap\gamma} m_{l1}(\beta) m_{l2}(\gamma)}{1 - \sum_{\beta\cap\gamma=\emptyset} m_{l1}(\beta) m_{l2}(\gamma)}, \tag{5}$$

which differs form the classical combination [3] in interpretation of α. The α fuzzy set is represented by intersections of fuzzy sets β and γ:

$$\mu_\alpha(x) = \mu_\beta(x) \wedge \mu_\gamma(x), \tag{6}$$

If some of these intersections are empty (i.e. results in zero membership), then a focal element can be eliminated by its zero BPA value. This means that a symptom is not considered for the $l-th$ diagnosis, but may remain significant for other diagnoses. The combination (5) can be done for two sources of information, but also for one training database.

Let us call the 'primary training' a determination of membership functions and BPAs by means of training data. After primary training classification of the training data is possible. If knowledge is perfect, the classification should be correct or very few misclassified cases should appear. In the opposite case, it is possible to create membership functions and BPAs separately for correctly classified and misclassified cases and next to combine them according to (5) and (6). If membership functions of the both kinds of cases have nothing in common, then a variable is eliminated from reasoning. As the example given below shows, such an elimination can be beneficial for inference.

3 Example

The iris database [1] was chosen to illustrate the method. This benchmark database does not particularly concern diagnosis, but it is well known, easy to handle and clearly illustrates inconsistent information represented by some variables, i.e. variables that do not differentiate classes. For these reasons it was chosen for the example. This database include small number of variables (4) and three classes (let us denote them D_1, D_2 and D_3), among them the first is easy to recognize [2]. Each class is represented by 50 cases. In the present example classification for this database is based on only two out of four variables, i.e. v_3 and v_4, which are correlated and more suitable to differentiate between classes 2 and 3. Only three focal elements for each diagnosis are created: $\mathbf{A_l}$ =$\{\alpha_{l1}, \alpha_{l2}, \alpha_{l3}\}$, $\alpha_{l1} \equiv$ 'v_3 is A_{l1}', $\alpha_{l2} \equiv$ 'v_4 is A_{l2}', $\alpha_{l3} \equiv$ 'v_3 is A_{l1} and v_4 is A_{l2}', $l = 1, 2, 3$. Classification in this problem is organized as follows. If for a case $Bel(D_1) > Bel(D_2)$ and $Bel(D_1) > Bel(D_3)$ than the case is classified as D_1, if $Bel(D_3) > Bel(D_1)$ and $Bel(D_3) > Bel(D_2)$ then conclusion is D_3. If the case is neither D_1 nor D_3 then it is D_2.

First, during primary training, membership function shapes were tested together with the BPA calculation. Since the latter required the choice of the η_{BPA} and for the classification η_T is necessary, the both thresholds were changed: η_{BPA} with 0.05 step and $\eta_T = \tau_i(\eta_{BPA})$, where τ_i were truth functions [6]: linear, square, third-power and third-root. In this way the most suitable value of the δ coefficient that determines the membership function shape was found. Simultaneously with determining the functions, the BPAs were calculated. The BPA values for single focal elements, i.e. $m_l(\alpha_{l1})$ and $m_l(\alpha_{l2})$ were multiplied by 2 to increase their significance in comparison to the complex focal element $m_l(\alpha_{l3})$, $l = 1, 2, 3$.

The primary training resulted in 4.67% of global error, i.e. overall error for three classes. The first class was correctly determined, the 6.00% error was obtained for the second class and 8.00% – for the third class. The most suitable thresholds were $\eta_{BPA} = \eta_T = 0.45$. The coefficient $\delta = -0.95$ which resulted in triangular functions. In Fig.1 diagrams *a)* and *b)* illustrate membership functions obtained during the primary training. Membership functions look reasonable, but the classification error is big.

Afterwards, correctly classified and misclassified cases were gathered in two sub-bases of training data. Membership functions and BPAs were again found

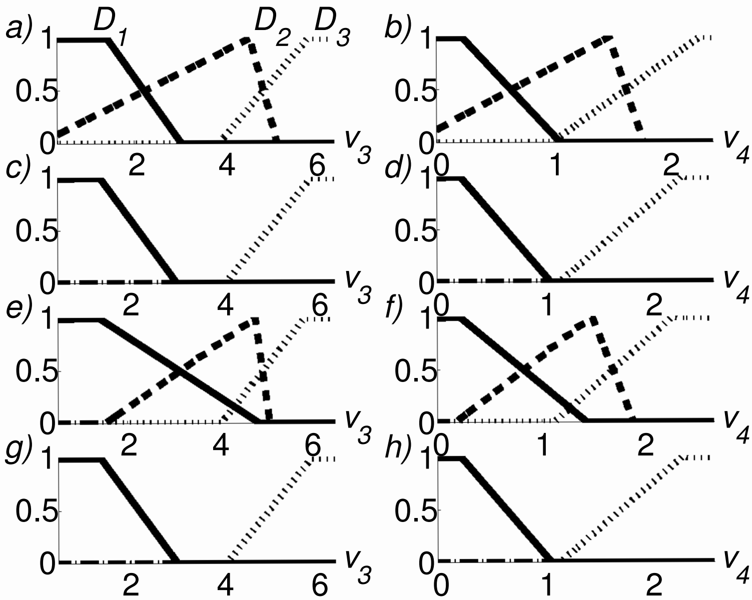

Figure 1: Membership functions obtained for variables v_3 and v_4 of iris data.

for the both sub-bases and next they were combined. The most suitable values for thresholds were searched beginning with their values obtained in the primary training and with the 0.02 step. Threshold were changed independently and the truth functions were not used anymore. The primary training should make the right threshold search quicker, which proved true with the inference threshold. The $\eta_T = 0.42$ was not much different from the threshold for primary training, but finding the appropriate $\eta_{BPA} = 0.98$ required more computations. In Fig.1 diagrams of membership functions obtained for each of the sub-databases are provided: functions resulted from correctly classified cases are in subplots *c)* and *d)*, while from misclassified cases are in *e)* and *f)*. The diagrams *c)* and *d)* show that the second class (D_2) was not recognized because there were no cases to construct membership functions for it. On the contrary, it was represented by misclassified cases, since for these cases all membership functions are built. The BPA values for correctly classified cases equal:

for class 1: $m_{cor1}(\alpha_1) = 0.35$, $m_{cor1}(\alpha_2) = 0.52$, $m_{cor1}(\alpha_3) = 0.13$,
for class 2: $m_{cor2}(\alpha_1) = 0$, $m_{cor2}(\alpha_2) = 0$, $m_{cor2}(\alpha_3) = 0$,
for class 3: $m_{cor3}(\alpha_1) = 0.49$, $m_{cor3}(\alpha_2) = 0.43$, $m_{cor3}(\alpha_3) = 0.07$,

while for misclassified cases $m_{misj}(\alpha_i) = m_{corj}(\alpha_i)$, $i = 1, 2, 3$, $j = 1, 3$ and:

for class 2: $m_{mis2}(\alpha_1) = 0.62$, $m_{mis2}(\alpha_2) = 0.33$, $m_{mis2}(\alpha_3) = 0.04$,

The diagrams *g)* and *h)* illustrate intersections of membership functions for the both sub-databases. Obviously, the membership functions for the second class do not exist again. This causes that the second class is 'the rest' of classification. The BPA values for the second class are equal to zero also as the result of combination (5) of greater than zero values of m_{mis2} with zero values of m_{cor2}. Yet, the assumption of classification to the second class each case that neither belongs to the first nor to the third class, results in errorless classification of this class. One third-class case is also classified as 'the rest', which results in 2% error for this class. The first class is correctly classified. All in all, for thresholds $\eta_{BPA} = 0.98$ and $\eta_T = 0.42$ classification error is 0.02 for the third class, the first and the second class are correctly classified. This means that the overall error for the three classes is equal $0.02/3 \approx 0.007$, which is smaller than in references with similar rules [2].
This simple example illustrates several advantages of the proposed method. Similar rules were constructed for the three classes (three diagnoses) and yet for one of the diagnoses rules were eliminated. The approach 'two classes and the rest' is known from references [2] and makes it possible to simplify rules. Rules that were created before training remained untouched during primary training and the combination, though their membership functions along with their weights were changed. Thus, if linguistic labels are attached to premises of the rules, they can be modeled by different membership functions, without changing their semantic meaning.

4 Conclusions

The proposed method makes it possible to assign membership functions and weights to rules that are IF-THEN heuristics. The membership functions represent fuzzy focal elements and weights are values of the basic probability assignment defined in the formalism of the Dempster-Shafer theory of evidence. Thus, this theory is extended for fuzzy focal elements. In the present method it is possible to attach various membership functions and weights to the same heuristic rules. The functions and the weights can be found in different manners and can be changed according to diagnostic circumstances. Particularly, they can be determined by means of training data. If some rules are not applicable to a training data set, they are eliminated because of zero membership functions in their premises or zero value of the basic probability. Simultaneously, values of the belief measures calculated for diagnostic hypotheses do not decrease. Therefore, interpretation of conclusions is always easy for a human diagnostician. The rule elimination can be performed after the primary step of partly successful classification by means of combining basic probability assignments for correctly and wrongly classified cases. Then, variables that disturb the classification can be removed and results can be improved.

References

[1] Asuncion A., Newman D.J., UCI Machine Learning Repository [http://www.ics.uci.edu/~mlearn/MLRepository.html], Irvine, CA: University of California, School of Information and Computer Science, 2007

[2] Duch W., Adamczak R., Grabczewski K., A new methodology of extraction, optimization and application of crisp and fuzzy logical rules, IEEE Transactions on Neural Networks, 11, 2000, 1–31.

[3] Kacprzyk J., Fedrizzi M. (eds.), Advances in Dempster-Shafer Theory of Evidence, J. Wiley, New York, 1994

[4] Nauck D., Kruse R., Obtaining interpretable fuzzy classification rules from medical data, Artificial Intelligence in Medicine, 16, 1999, 149–169.

[5] Rutkowska D., IF-THEN rules in neural networks for classification, Proc. Int. Conf. on Computational Intelligence for Modelling, Control and Automation, and Int. Conf. Intelligent Agents, Web Tchnologies and Internet Commerce, 2, IEEE Computer Society, Washington, USA, 2005, 776–780.

[6] Straszecka, E., Combining uncertainty and imprecision in models of medical diagnosis, Information Sciences, 176, 2006, 3026–3059

The papers presented in this Volume 2 constitute a collection of contributions, both of a foundational and applied type, by both well-known experts and young researchers in various fields of broadly perceived intelligent systems.

It may be viewed as a result of fruitful discussions held during the Eighth International Workshop on Intuitionistic Fuzzy Sets and Generalized Nets (IWIFSGN-2009) organized in Warsaw on October 16, 2009 by the Systems Research Institute, Polish Academy of Sciences, in Warsaw, Poland, Centre for Biomedical Engineering, Bulgarian Academy of Sciences in Sofia, Bulgaria, and WIT – Warsaw School of Information Technology in Warsaw, Poland, and co-organized by: the Matej Bel University, Banska Bistrica, Slovakia, Universidad Publica de Navarra, Pamplona, Spain, Universidade de Tras-Os-Montes e Alto Douro, Vila Real, Portugal, and the University of Westminster, Harrow, UK:

http://www.ibspan.waw.pl/ifs2009

The Eighth International Workshop on Intuitionistic Fuzzy Sets and Generalized Nets (IWIFSGN-2009) has been meant to commence a new series of scientific events primarily focused on new developments in foundations and applications of intuitionistic fuzzy sets and generalized nets pioneered by Professor Krassimir T. Atanassov. Moreover, other topics related to broadly perceived representation and processing of uncertain and imprecise information and intelligent systems are discussed.

We hope that a collection of main contributions presented at the Workshop, completed with many papers by leading experts who have not been able to participate, will provide a source of much needed information on recent trends in the topics considered.

ISBN-13 9788389475305
ISBN 838947530-8
9 788389 475305